CLAIMS

1. A method for processing information contained in a body of given original data, the method comprising the steps of:
- pointing out at least two exemplary cases,
- comparing the at least two exemplary cases to each other for finding congruent parts from them,
- as a result of the comparing, generating a regular expression, which describes the appearance of congruent parts in the at least two exemplary cases,
- on the basis of the generated regular expression, generating a set of rules for extracting data of a desired kind, and
- extracting data areas from the original data according to the generated set of rules.

Sub A3

2. A method according to claim 1, comprising the step of modifying the extracted data areas to be uniform in format.

3. A method according to claim 1, wherein the at least two exemplary cases that are pointed out each have a structure and a content, and the structure of the exemplary cases is identical, but the content is different.

4. A method according to claim 1, wherein the exemplary cases are pointed out from the original data.

5. A method according to claim 1, wherein the regular expression comprises the congruent parts and wildcard expressions, which correspond to matter to be extracted.

6. A method according to claim 1, wherein the set of rules generated on the basis of the regular expression is stored for further usage.

7. A method according to claim 1, comprising the step of tokenizing the chosen exemplary cases prior to the processing proper thereof, by replacing certain elements of the exemplary cases by corresponding data structures, which contain an identifier, such as a type characteristic, or a name, as well as a data content of said element.

8. A method according to claim 7, wherein between the at least two exemplary cases pointed out, there is at least one identical element, the counterpart whereof in the treatment of the exemplary cases is a given token.

SubA3

~~9. A method according to claim 7, wherein in order to generate a set of rules, the method comprises the steps of:~~

- ~~marking the longest of the selected, tokenized examples as a regular expression,~~
- ~~marking the next longest of the selected, tokenized examples as an exemplary expression and~~
- ~~comparing the regular expression with the exemplary expression of the moment in question.~~

~~10. A method according to claim 9, wherein the regular expression and the exemplary expression of the moment in question are compared by means of a given reference algorithm that returns an edit script.~~

~~11. A method according to claim 10, wherein the regular expression and the exemplary expression of the moment are compared by means of a reference algorithm that returns the shortest possible edit script.~~

12. A method according to claim 10, wherein in order to generate a set of rules, the regular expression is modified according to the edit information contained in the edit script.

13. A method according to claim 9, wherein the created regular expression constitutes a set of rules by itself.

14. A method according to claim 1, wherein by means of the generated set of rules, from the original data there are extracted elements according to the exemplary cases.

15. An arrangement for processing information contained in a given original data, comprising:

- means for pointing out at least two exemplary cases,
- means for comparing the at least two exemplary cases to each other for finding congruent parts from them,
- means for generating, as a result of the comparing, a regular expression, which describes the appearance of congruent parts in the at least two exemplary cases,
- means for generating a set of rules on the basis of the generated regular expression, in order to extract desired information, and
- means for extracting data areas from the original data according to the generated rules.

Sub A4

16. An arrangement according to claim 15, comprising means for modifying the extracted elements to be uniform in format.

17. An arrangement according to claim 15, wherein in order to point out exemplary cases, the arrangement is provided with pointers to character strings.

18. An arrangement according to claim 15, comprising means for tokenizing the examples pointed out by replacing given elements of the exemplary cases by corresponding data structures that contain a type characteristic or a name as well as a data content of said element.

19. An arrangement according to claim 18, wherein the arrangement includes means for processing tokenized data.

20. An arrangement according to claim 15, wherein the arrangement includes means for generating a set of rules according to a created regular expression.

21. An arrangement according to claim 15, wherein the means for generating the set of rules including a program component created especially for this purpose, which program component is different from the program component that is meant for extracting data areas by using the generated set of rules.

22. A computer program element comprising: computer program code means to make the computer execute a procedure that comprises the steps of:
- pointing out at least two exemplary cases,
- comparing the at least two exemplary cases to each other for finding congruent parts from them,
- as a result of the comparing, generating a regular expression, which describes the appearance of congruent parts in the at least two exemplary cases,
- on the basis of the generated regular expression, generating a set of rules for extracting data of a desired kind, and
- extracting data areas from the original data according to the generated set of rules.

23. A computer readable medium, having a program recorded thereon, where the program is to make the computer execute a procedure that comprises the steps of:
- pointing out at least two exemplary cases,
- comparing the at least two exemplary cases to each other for finding congruent parts from them,
- as a result of the comparing, generating a regular expression, which describes the appearance of congruent parts in the at least two exemplary cases,

- on the basis of the generated regular expression, generating a set of rules for extracting data of a desired kind, and
- extracting data areas from the original data according to the generated set of rules.

24. A computer program product stored on a computer usable medium, comprising: computer readable program means for causing a computer to execute a procedure that comprises the steps of:

- pointing out at least two exemplary cases,
- comparing the at least two exemplary cases to each other for finding congruent parts from them,
- as a result of the comparing, generating a regular expression, which describes the appearance of congruent parts in the at least two exemplary cases,
- on the basis of the generated regular expression, generating a set of rules for extracting data of a desired kind, and
- extracting data areas from the original data according to the generated set of rules.